Performance Modelling of Patient Flow Scheduling Through a Formal Method

陈潇 [1*], Nigel THOMAS[2], 丁杰 [3]
(1. School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu, China.
2. School of Computing Science, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK.
3. School of Information Engineering, Yangzhou University, Yangzhou, Jiangsu, China.)
*Email address: xiaochen@ujs.edu.cn
Received Date: August 30, 2016.

Abstract: **Smart environment is being used in many areas to deliver more services to individuals in a physical space, such as a hospital. In the UK, the National Health Service (NHS) provides free and high quality healthcare service for all residents. Smart hospital environment (SHE) is able to support NHS and provide more convenience. Patient flow scheduling is a crucial section in a smart hospital environment. SHE aims to provide a smart environment in the hospital to facilitate individual experience and improve the quality of healthcare service. First of all, this paper investigates a real world patient flow scenario of a hospital in the UK and models a general scheduling scheme based on the scenario using a compositional formal approach, i.e. Performance Evaluation Process Algebra (PEPA). This scheduling scheme uses a easy-implemented solution (the grouping scheme) to reduce the waiting queue in the hospital. Secondly, fluid flow analysis is used for the performance analysis by generating a set of ordinary differential equations (ODEs) in terms of the PEPA model.**

Keywords: scheduling, performance valuation, patient flow, PEPA, ODEs

CLC Number: TP302, TP391

0 Introduction

Nowadays, emerging technology, growing pressure to improve quality, heightening patient expectations, shorter patient stays, and aging population have made the management of healthcare resources even more critical. Healthcare in the UK is mainly provided by the NHS which provides healthcare to all residents freely. Thus, the government may concern with the utilization of healthcare resources. Hospitals may be interested in how many consultants should be employed in each department. For patients, their main concern is about how fast they can get their healthcare. All these issues refer to the capacity planning for a healthcare department and scheduling development for the patient flow.

The research is financially supported by National Nature Science Foundation of China with Grant No. 61502206, 61472343, Natural Science Foundation of Jiangsu Province of China with Grant No. BK20160543, BK20150523 and Open Project of Key Laboratory of Jiangsu Province with Grant No. BM20082061507.

With the development of computing technology especially the cyber physical system (CPS), it is possible to build a smart hospital environment to solve the issues mentioned before. Our previous research investigates an efficient capacity plan based on the scenario of the rheumatology department of a hospital in the UK. However, in previous research, we just considered the capacity planning issue in terms of the requirements rather than the scheduling issue. As a result, the waiting queue of patients will have a sharp increase in the hospital. To facilitate the experience of patients, a scheduling scheme is required to reduce the waiting queue in order to save the patients' time.

Gartner and Kolisch [1] explored the problem of planning the patient flows due to the scarce medical resources using a mathematical approach. In the research, two mixed-integer programs are developed to achieve a significant improvement of the contribution margin. Kortbeek et al. [2] investigated the performance of the appointment schedules that balance the waiting time for the unscheduled patients and the access time for the scheduled patients based on the iterative calculation of two designed models. In this research, numerical experiments are used to generate the comparison outcomes. From these studies, we can clearly see that most researches about the patient flow scheduling are conducted through a numerical approach. However, in this paper, we use a compositional formal method i.e. Performance Evaluation Process Algebra (PEPA) to model a scheduling scheme with a patient flow scenario.

This paper aims to explore a scheduling scheme for the patient flow model in order to reduce the waiting queue in the hospital. Such scheme is designed to cut the continuous incoming patient flow into different groups. Each group is scheduled in different time slots in a working day in order to avoid a large number of arrivals in the same time slot, such as the morning rush hour. Such grouping scheme is efficient to reduce the queue by scheduling patients in different time slots. Moreover, this scheme is easy to be implemented based on the current clinic appointment system. In this case, the patient flow scenario is a kind of Markov process. Thus, the underlying Markov model of patient flow can be created through the formal method PEPA.

The main contribution of this work can be highlighted as follows. An efficient scheduling scheme for the patient flow is proposed. A formal method is used to model a complex patient flow model with scheduling mechanism. The fluid flow approximation for the model analysis without a steady state is used.

1. Model Scenario

This research is based on a scenario of rheumatology department of Royal Hallamshire Hospital in Sheffield. Rheumatology department runs the service for a number of patients with a fixed procedure. The patient flow information is collected from the staff of rheumatology department in an interview. In addition, a set of internal statistical figures observed from the daily records are also collected from the department. The real situation in rheumatology department is depicted in Figure 1.

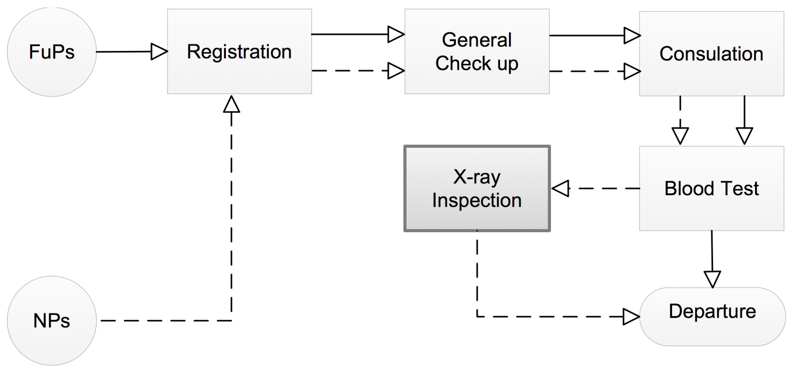

Figure 1. Patient Flow of Rheumatology Department

As displayed in Figure 1, patients can be generally grouped in two classes in terms of the research objects, namely follow-up patients (FuPs) and new patients (NPs). These two patient flows are indicated in Figure 1 with a solid line for the FuPs and a dotted line for the NPs. Each rectangle represents a step point of the whole process. The both groups of patients must register for an appointment (registration) with the department first, and then each patient has a general check up before meeting the consultants. After seeing the consultants, FuPs go for a blood test and leave the department. However, NPs must go to another department for an X-ray inspection after the blood test, and then leave the department.

2. Modelling Approach

Now we introduce a novel approach for performance modelling. It is a type of stochastic process algebra, PEPA. As a high-level modelling paradigm for continuous time Markov chains (CTMC), a compositional structure is inherent in the PEPA language [6]. This compositionality may be exploited to reduce the state space of the CTMC [5]. Furthermore, this technique takes advantages of symmetries within the system, and may be formally defined based on the model's PEPA description. Thus, a PEPA model with the underlying CTMC can be applied to define a Markov Reward Process (MRP) from which performance evaluation can be derived [6].

Queuing network provides compositionality but not formality; stochastic Petri net offers formality but not compositionality. Neither give abstraction mechanism. Finally, both complete content and organizational editing before formatting.

2.1 Syntax of PEPA

PEPA has only four combinators, which are *prefix*, *choice*, *cooperation* and *hiding* [5].

Prefix: It is the basic building block of a sequential component: the process $(a, r).P$ performs action a at rate r and then evolves to component P.

Choice: It generates a competition between two or more potential processes: $(a, r).P + (b, s).Q$ represents that either action a or b wins the race at the rate r or s and then behaves as P or Q respectively.

Cooperation: Its operator joins two processes together, in which these two processes may share some actions: the process $P \bowtie_L Q,\ L = \{a, b\}$ denotes that components *P* and *Q* must cooperate with the shared actions *a* and *b*. $\bowtie$ represents a cooperation as well as a set of shared actions specified in L. Any other action is performed independently. Additionally, $P \parallel Q$ in PEPA syntax means $P \bowtie_L Q$, $L = \emptyset$.

Hiding: the process $P|\{a\}$ hides the action *a* from view and prevents other processes from joining with it.

Constant: As assumed, a countable set of constants is defined. The constants are components with the meaning of defining an equation such as $A \stackrel{\text{def}}{=} P$ which gives the constant *A* the behaviour of the component *P*. This is the way of assigning names for behaviour components.

The syntax of PEPA in describing the above processes is given as:

$P := (a, \lambda).P \mid P + Q \mid P \bowtie_L Q \mid P|Q \mid A$

This PEPA statement involves all four combinators mentioned in the previous paragraph. The last part of this statement $P := A$ is to identify component *P* with *A*. When the rate of the action is passive, we use the symbol ⊤. More details of PEPA syntax and examples can be found in Ref. [5].

2.2 Fluid Flow Approximation

PEPA language offers compositional function for creating models towards large-scale systems. Meanwhile, a novel performance analysis technique, i.e. fluid flow approximation, is provided for large-scale models of PEPA.

Just as most discrete state-based modelling formalities, process algebra easily suffers from the failure due to the generation of extremely large state space making the numerical solution via linear algebra costly or even intractable [16]. Fluid flow approximation generates a set of coupled ordinary differential equations (ODEs) underlying continuous time Markov chains. This approach successfully avoids the state space explosion in the analysis via exploring ODEs.

3. Patient Flow Model

This section briefly introduces the previous work of investigating the efficient capacity plan based on the patient flow displayed in Figure 1. Analysis is conducted based to the PEPA using the fluid flow approximation and the stochastic simulation for the validation.

3.1 Initialization of the Patient Flow Model

In this research, the cycle time at each service component is the statistical figure obtained from the rheumatology department (see Table 1).

Table 1. Statistical Parameters Used in the Model

Component	Cycle Time /s	Service Rate (persons/s)
Registration	81	44
General Check up	106	34
FuPs	15	4
NPs	30	2
Blood test	252	14
Departure	243	14

According to the cycle time t_C, the service rate μ, that is the number of patients served per hour, at each component in the workflow can be calculated in terms of the formula (1). In all situations, the time unit is unified as hour.

$$\mu = \frac{1}{t_c} \qquad (1)$$

In addition to the statistic parameters, a set of hypothetic parameters are used in the model. On each working day, 20 NPs and 80 FuPs must be served. Thus, there are, on average, 2.5 NPs and 10 FuPs arriving in the department each hour.

In terms of the cycle time and Takt time (Takt time is the average unit production time needed to meet customer demand), the instance number of each component can be determined. For example, the cycle time for consulting a NP is 30 minutes; however, the Takt time for consulting a NP needs 24 minutes. Thus, at least 2 consultants for NPs are required to in-sure that all NPs can be severed on time. In the same way, the number of consultants serving the FuPs should be 3 at least. Hence, according to the statistic data, the instance number of each component used in the model is shown in Table 2.

Table 2. Number of Components Used in the Model

Component	Instance number
NPs	20
FuPs	80
Registration	1
General Check up	1
NPs Consultant	2
FuPs Consultant	3
Blood test	2
X-ray inspection	2
Departure	2

3.2 Model Analysis

The analysis is generated based on PEPA with two different analyzing techniques: stochastic simulation and fluid flow approximation. The models are run on the Eclipse platform with PEPA plug-in. Fluid flow approximation is carried out with the adaptive step-size 5-order Dormand Prince ODE solver from time point 0 to 10, with 1000 data points and a step size of 10^{-3} time units. The relative error and absolute error equal to 10^{-4}time units. All parameters used in validation are noted in Table 1 and Table 2.

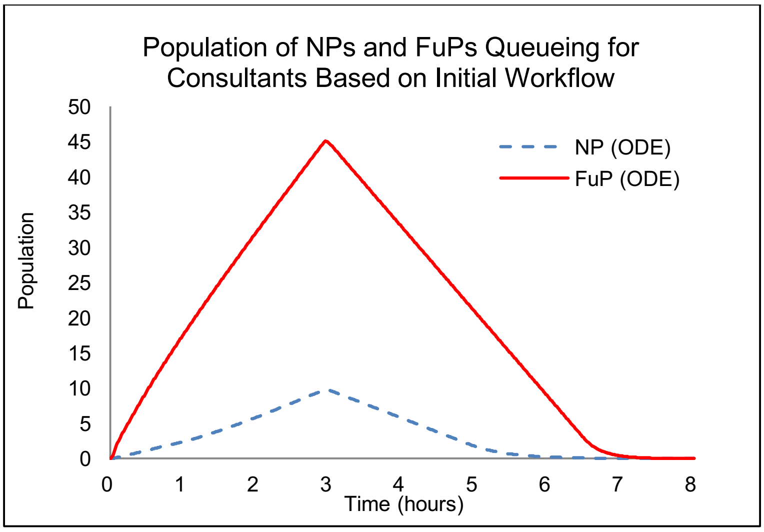

Figure 2. Population of NPs and FuPs Queueing for Consultants Based on Initial Workflow

Figure 2 displays the transient queue length of the NPs and the FuPs patients waiting for the consultants in the department. As shown in the figure, two kinds of patients come to the department during the first three hours and then the queue increases to a peak value. This is because all patients start going to the hospital at the same start time in terms of the model definition. Thus, these patients will suffer from a crash in the department which causes a long waiting queue.

The long waiting queue becomes a series of problems in the hospital department. There are two possible ways to reduce the waiting queue: the first is to refine the patient flow to improve the work efficiency; the second is to apply a scheduling policy for the administration process to control the number of incoming patients at each time slot. In the previous work, the patient flow has been evolved through changing the order of some activities in the flow. Next section will give a short demonstration of this improvement.

3.3 Model Evolution

This section demonstrates an evolved patient flow of the rheumatology department. In contrast to the original scenario, the flow of the NPs is changed to a new order, which is to bring forward the X-ray inspection and the blood test before the test, and the FuPs flow remains the same. The evolved patient flow is shown in Figure 3.

In Figure 3, the dotted line clearly depicts the flow of the NPs in the rheumatology department. As the X-ray inspection must be handled in another department, it is more efficient for the NPs to complete X-ray inspection before entering the rheumatology department rather than leaving for X-ray inspection halfway and coming back. Furthermore, bringing forward the X-ray inspection and the blood test for the NPs is to schedule more consultants from the NPs to the FuPs, when the NPs are leaving for the X-ray inspection and doing the blood test.

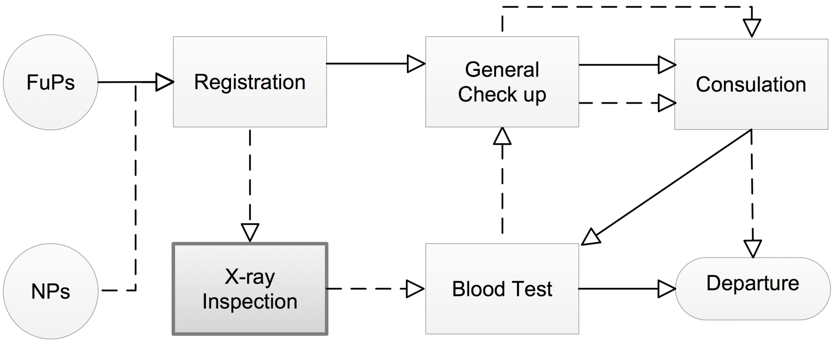

Figure 3. Evolved Patient Flow of Rheumatology Department

Thus, in the evolved model, consultants previously working for the NPs now must work for all patients. When the NPs leave for X-ray inspection or carry out the blood test, these consultants serve the FuPs; once the NPs come back, these consultants need serve the NPs first. The analysis of this evolved model is also based on the fluid flow approximation. Conditions of analyzing methods and parameters of the model remain the same, as shown in Table 1 and 2.

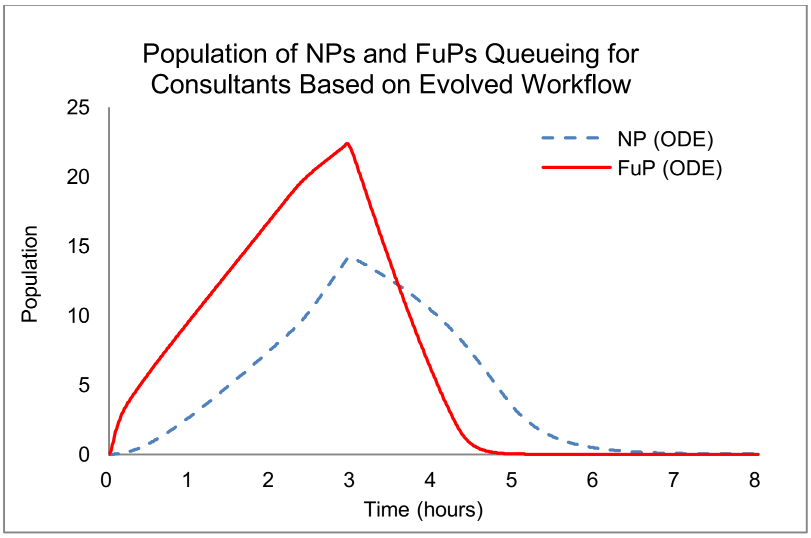

Figure 4. Population of NPs and FuPs Queueing for Consultants Based on Evolved Workflow

Figure 4 indicates the number of patients queuing for consulting service without using any scheduling policy. In contrast to Figure 2, the peak queue length of the FuPs is reduced from 48 to 23, because some consultants for the NPs come to serve the FuPs when they are free. However, the peak queue length of the NPs has a little increase from 10 to 14. Both lines still have dramatic increase within the first three hours and reach a peak value, which is quite similar to the situation in Figure 2.

Hence, it is clear that even if the patient flow is refined and the number of consultants is increased, the queue length still cannot be effectively reduced. Thus, just refining the workflow is not enough to reduce the queue. In the next section, a grouping based scheduling scheme will be introduced to solve the long queue problem.

4. Model of Scheduling Scheme

The section aims to investigate an alternative grouping scheme based on the evolved model, and generate the performance analysis through the fluid flow approximation. The previous modelling is processed with the assumption that all consultants work continuously without rest until completing all tasks. However, in the real situation, the consultants cannot work the whole day without rest. Thus, the consultants must have several breaks during a working day. Then, the patients should be scheduled into several groups so that each group comes to the department in a different time slot; the next group always follows the previous. This section explores a discontinuous situation based on the evolved patient flow model.

4.1 Scenario Description and Model Construction

To model the discontinuous patient flow, each patient group (NPs or the FuPs) must be split into four equivalent small sub-groups. In the model, each sub-group is registered to enter the department after the same specified time period. For example, the time period is assumed to be 2 hours. Thus, Group 1 starts entering the department from the time point zero; Group 2 starts from the time point 2; similarly, Group 3 and Group 4 starts from the time point 4 and the time point 6 respectively. Once the patient group starts, they will be scheduled under the static policy first, and then follow the evolved flow until departure.

To achieve this scenario, first of all, the evolved model type must be changed to the type with four sub-groups, which is displayed as follows:

$$NPs1 \stackrel{\text{def}}{=} (arriveNew1, r_{arriveNew1}).NPs1_{reg}$$

$$NPs1_{reg} \stackrel{\text{def}}{=} (register,\ r_{register}).NPs1_{xray}$$

$$NPs1_{xray} \stackrel{\text{def}}{=} (xray, r_{xray}).NPs1_{blood}$$

$$NPs1_{blood} \stackrel{\text{def}}{=} (blood, r_{blood}).NPs1_{test}$$

$$NPs1_{test} \stackrel{\text{def}}{=} (test, r_{test}).NPs1_{con}$$

$$NPs1_{con} \stackrel{\text{def}}{=} (newCon, r_{newCon}).NPs1_{depart}$$

$$NPs1_{depart} \stackrel{\text{def}}{=} (depart, r_{depart}).Stop$$

The above model is created to model the process of Group 1 involving the NPs. It defines a series of activities of NPs with the following action types arriving ($arriveNew1$), registration($arriveNew1$), X-ray inspection ($xray$), blood test ($blood$), general check up ($test$), consultation ($newCon$) and departure ($depart$), as well as their corresponding action rates. From the model details, the patient flow is completely the same as the previous evolved model. For the other groups, they all have the same process.

Although the above model represents four groups of patients, the key point is to make each sub-group starting in different time slots. Therefore, a function rate must be used for the arrival rate of each group, which is expressed as follows:

$$\lambda = \begin{cases} C, & t > T \\ 0, & t \leq T \end{cases},$$

Where, λ is the arrival rate, t is the transient time in the model run; C is a constant value; and T is the specified timeout. This formula means that the arrival rate remains to be zero until the time point T; and after T, the rate is C. The value of T is chosen based on the number of groups and the final complete time. The function accomplishes the process that each patient sub-group starts from different time.

4.2 Model Analysis

In the analysis, it is assumed that the time span to start each sub-group is 2 hours. The fluid flow approximation will be processed with MatLab in the following analysis. Here, ODE45 in MatLab is adopted to solve ODEs; time span is from point 0 to 11; step size is 0.01 time units.

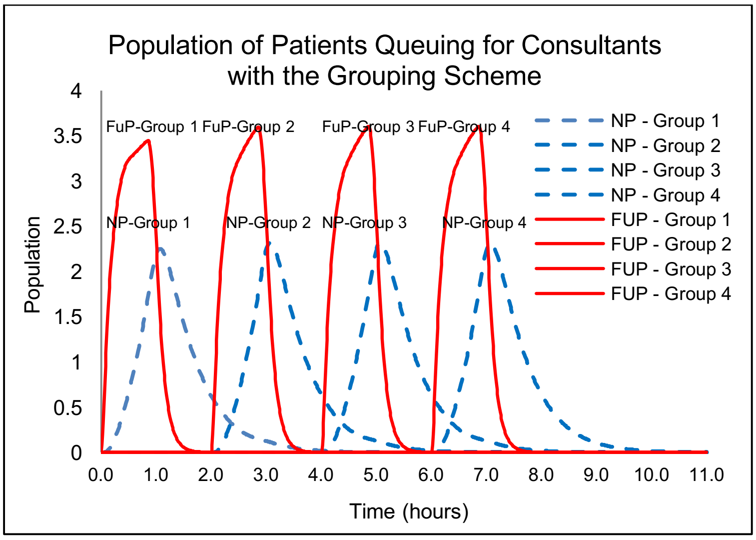

Figure 5. Population of Patients Queuing for Consultants with the Grouping Scheme

Figure 5 clearly shows the queue length for consulting of each patient group. The dotted lines represent the population of the NPs, and the dashed lines stand for the population of the FuPs. From the figure, it is easy to distinguish four groups of patients according to the waves. The subsequent sub-groups always start at the end of the previous group. Owing to a static factor and the updated workflow used in the model, the queue length of the NPs is just around 3.5 at the peak, and the queue of the FuPs is about 2.2 for the peak value. The population waiting for consultants fluctuates like waves. The wave peak means that the more patients are waiting for service, and then the consultants are busy. However, at the end of each sub-group, there are fewer patients queuing for service, so, the consultants have some free time for a break.

Compared with the FuPs' population, the NPs' population has a slower increase after the start in each sub-group. This is because the workflow, in which the NPs need complete x-ray inspection and blood test first before seeing the consultants. It is one reason causing the NPs late completion. Another reason is the number of consultants serving the NPs. In this case, there are 3 consultants serving both the NPs and the FuPs. However, there is no consultant serving the NPs only. Thus, to solve this problem, one way is to add more consultants who serve the NPs only. The third possible reason is the time span setting to start each sub-group. Here, the time interval of 2 hours is used as a slot. To reduce the total time span, each time span for an individual sub-group can be set less than 2 hours, for example, 1.5 hours.

5. Conclusion

This paper aims to develop a simple grouping scheduling scheme for the rheumatology department based on the previous capacity model in order to solve the long waiting queue problem. The main technique used in modelling is PEPA based formal method and the fluid flow approximation for model solving and analysis. According to the analysis results, the grouping scheduling scheme is able to effectively reduce the waiting queue, and this scheme is easier to implement on the current appointment system than other scheduling policies which may need extra physical devices and applications to support.

Acknowledgment

The research is financially supported by National Nature Science Foundation of China with Grant No. 61502206, 61472343, Natural Science Foundation of Jiangsu Province of China with Grant No. BK20160543, BK20150523 and Open Project of Key Laboratory of Jiangsu Province with Grant No. BM20082061507.

REFERENCES

[1] D. Gartner and R. Kolisch, "Scheduling the hospital-wide flow of elective patients", European Journal of Operational Research, Vol.233, Issue 3, pp.689-699, March 2014.

[2] N. Kortbeek, M. E. Zonderland, A. Braaksma, I. M. H. Vliegen, R. J. Boucherie, N. Litvak, and E. W. Hans, "Designing cyclic appointment schedules for outpatient clinics with scheduled and unscheduled patient arrivals", Performance Evaluation, Vol. 80, SI: Service Science of Queues, pp.5-26, October 2014.

[3] J. Tang, C. Yan, P. Cao, ”Appointment scheduling algorithm considering routine and urgent patients”, Expert Systems with Applications, Vol. 41, Issue 10, pp.4529-4541, August 2014.

[4] R. R. Chen and L. W. Robinson, ”Sequencing and Scheduling Appointments with Potential Call-In Patients”, Production and Operations Management, Vol. 23, No. 9, pp.1522-1538, September 2014.

[5] J. Hillston, ”A compositional approach to performance modelling”, Cambridge University Press, 1996.

[6] J. Hillston, “Fluid Flow Approximation of PEPA Models”, IEEE Proceedings of the Second International Conference on the Quantitative Evaluation of Systems, 2005.

[7] X. Chen, L. Wang, J. Ding and N. Thomas，“Patient Flow Scheduling and Capacity Planning in a Smart Hospital Environment”， IEEE Access，Vol. 4 , Issue 1，pp. 1-14, 2016.

[8] X. Chen and L. Wang, “Exploring Trusted Data Dissemination in a Vehicular Social Network with a Formal Compositional Approach”, 2016 IEEE 40th Annual Computer Software and Applications Conference (2016 COMPSAC), Atlanta, US, 2016.